

Corruption in public service delivery: an experimental analysis

Barr, Abigail; Lindelow, Magnus; Serneels, Pieter

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:

Barr, A., Lindelow, M., & Serneels, P. (2009). Corruption in public service delivery: an experimental analysis. *Journal of Economic Behavior & Organization*, 72(1), 225-239. <https://doi.org/10.1016/j.jebo.2009.07.006>

Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu>. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under the "PEER Licence Agreement". For more Information regarding the PEER-project see: <http://www.peerproject.eu>. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

Accepted Manuscript

Title: Corruption in Public Service delivery: an experimental Analysis

Authors: Abigail Barr, Magnus Lindelow, Pieter Serneels

PII: S0167-2681(09)00183-8
DOI: doi:10.1016/j.jebo.2009.07.006
Reference: JEBO 2420

To appear in: *Journal of Economic Behavior & Organization*

Received date: 4-3-2008
Revised date: 16-7-2009
Accepted date: 17-7-2009

Please cite this article as: Barr, A., Lindelow, M., Serneels, P., Corruption in Public Service delivery: an experimental Analysis, *Journal of Economic Behavior and Organization* (2008), doi:10.1016/j.jebo.2009.07.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



CORRUPTION IN PUBLIC SERVICE DELIVERY: AN EXPERIMENTAL ANALYSIS

Abigail Barr
University of Oxford

Magnus Lindelow
The World Bank

Pieter Serneels¹
University of East Anglia

December 2008

Abstract²: To improve our understanding of corruption in service delivery, we use a newly designed game that allows us to investigate the effects of the institutional environment on the behavior of service providers and their monitors. We focus on the effect of four different factors: whether monitors are accountable to the service recipients, the degree of observability of service providers' effort, the providers' wages and the providers' professional norms. In accordance with theory, we find that service providers perform better when monitors are elected by service recipients and when their effort is more easily observed. However, there is only weak evidence that service providers perform better when paid more. Monitors are more vigilant when elected and when service providers are paid more. Playing the game with Ethiopian nursing students, we also find that those with greater exposure to the Ethiopian public health sector perform less well, either as provider or as monitor, when the experiment is framed as a public health provision scenario, suggesting that experience and norms affect behavior.

¹ Corresponding author, email: p.serneels@uea.ac.uk

² We would like to thank the attendants of the North American ESA conference, as well as two anonymous referees for useful comments, and Omar Azfar and William Nelson for comments on experimental design. We would also like to thank Ritva Reinikka and Agnes Soucat for helpful discussions and support, and Beliyu Astatkie and Sitotaw Enyew for assistance in implementation of the experiments. Financial support from The World Bank and cooperation from the Ministry of Health in Ethiopia are greatly acknowledged. The findings, interpretations and conclusions of this paper are those of the authors, and should not be attributed to the World Bank, its executive directors, or the countries they represent.

1. Introduction

This paper investigates the impact of institutional design on individual behavior in a particular, frequently occurring situation. The situation involves a principal who contracts an agent to deliver a service, not to the principal, but to a third party. Because the contracting principal is not the recipient of the service, he is not automatically in possession of the information required to judge whether the agent is fulfilling the contract and this limits his ability to enforce that contract. In an effort to address this problem, the principal employs a monitor to collect the required information and report back or, in some contexts, process the information, pass judgment on the agent and act accordingly. However, this creates another problem as the relationship between the principal and the monitor displays features similar to the relationship between the principal and the original, service-supplying agent.

Situations fitting this characterization are commonplace. However, our particular interest is situations in which this type of principal-agent problem impacts public service delivery. Placing the government in the role of contracting principal, the vast majority, if not all, public servants can be likened to either the service provider or the monitor in the scenario described above.

With respect to both intent and approach, our endeavour has much in common with that of Ostrom et al. (1992). Our intent is to identify causal links between elements of institutional design and behavior. Our approach is to involve human subjects in an experiment involving a game that simulates the natural-world situation in which we are ultimately interested. However, while Ostrom et al. (1992) focused on the provision of public goods by societies, we focus on the provision of public services by agents employed to serve societies. To this end, we undertake an economic experiment involving a game specifically designed to mimic the situation described above. Within the context of that experiment we vary four factors, each of which has been identified as potentially important in the existing literature, and observe the resulting effects on service provider and monitor behavior. Three out of the four factors we vary are institutional. The first is whether the service recipients have any power to hold those charged with the duty of monitoring service providers accountable; the second is the ease with which the

performance of service providers can be monitored; and the third is service provider pay. The fourth factor we vary is the apparent salience to the experimental context of professional norms that may be referenced and/or adhered to by our subjects.

Our experimental subjects are Ethiopian nursing students, all destined to work in the Ethiopian public health sector which, like many public sectors throughout the developing world, is plagued by problems of absenteeism and the misappropriation or misuse of consumable resources and equipment by often highly demoralized public servants. It is this unusual choice of subjects that facilitates our exploration into the effect of professional norms on individual behavior. This choice of subjects may also increase the external validity of our results.

The paper has 6 sections. Following this introduction, section 2 explains our rationale for using an experiment to explore the impact of the four factors listed above on public service delivery. In section 3 we describe the game. Section 4 provides information about the Ethiopian context and the subjects participating in the experiments. In section 5 we present our results. And in section 6 we conclude.

2. An experimental approach

One of the standard frameworks used in the theoretical analysis of corruption is the principal-agent model. Using this model researchers have investigated the trade-offs between the expected costs and benefits of corrupt acts (Bardhan, 1997; Klitgaard, 1988; Rose-Ackerman, 1978). One of the findings resulting from this work is that lower wages increase the likelihood of corruption by reducing the cost associated with being caught (Becker and Stigler, 1974; Chand and Moene, 1999; Mookherjee and Png, 1995) and through adverse selection (Besley and McLaren, 1993).

Principal-agent models have also been used to explore the impact of monitoring and sanctioning on agent performance and of mutually beneficial, corrupt transactions between monitors and agents on social welfare (Becker and Stigler, 1974; Khalil and Lawarree, 1995; Mookherjee and Png, 1995). The role of community or service recipient involvement in monitoring has also been addressed, principally in

the literature on decentralization, where some have argued that community participation strengthens accountability and improves governance, while others have questioned this, pointing to the lack of capacity at the community-level to interpret and act upon available performance-related information (Banfield, 1979; Manor, 1999; Prud'homme, 1995).

However, even when the monitoring and sanctioning is conducted by public servants, capacity can be an issue as public service providers' outputs are often both multidimensional and intrinsically difficult to observe (Dixit, 1997, 2001; Le Grand, 2003; Wilson, 1989). And acknowledgement of this has given rise to a range of interventions aimed not at monitoring and enforcement but at aligning the incentives of principals and agents. Many of these interventions take the form of structural reforms designed to build up and enhance the effects of organizational commitment, social and moral norms and intrinsic motivations on individual performance.³ These factors, often collectively referred to as professional norms, have received particular attention in the health sector where monitoring is especially difficult (Arrow, 1963). Indeed, some (e.g., Frank, 2004) have argued that such norms are fundamental in constraining opportunistic behavior by health service providers and that this explains the prevalence of non-profit organizations in the health sector (Glaeser and Shleifer, 2001; Pauly, 1987).

While strands of the theoretical literature point to the potential importance of wages, monitoring arrangements, and professional norms, it has proven difficult to test predictions empirically. In the case of wages and monitoring, empirical investigations have, with few exceptions (Ablo and Reinikka, 1998; Fisman and Gatti, 2000; Svensson, 2003), been based on cross-country perception data. These analyses provide some support for the theoretical predictions relating to aspects of the legal system, public sector recruitment policies and salaries, and various dimensions and degrees of decentralization (for a review, see Andvig and Fjellstad (2001)), but have been criticised on account of the unsatisfactory measurement of corruption, the potential endogeneity of some determining factors, and problems of robustness. These

³ Intrinsic motivation can be defined as the motivation that drives an individual to conduct a task - whatever the external incentives (Deci, 1975; Frey, 1997; Lindenberg, 2001). Intrinsic motivation comes about through the internalization of norms and has been argued to be a stronger enforcement mechanism than external norms (Coleman, 1990).

problems are especially severe in the work on public servant pay. Thus, while Van Rijckeghem and Weder (2001) found that higher public servant salaries (relative to manufacturing wages) are associated with less corruption, neither Treisman (2000) nor Rauch and Evans (2000) were able to replicate the result using different proxies for the explanatory variable. Moreover, Rauch and Evans (2000) found that employment security and recruitment and promotion criteria may be more important.

Evidence relating to the role of intrinsic motivations and professional and social norms is even more limited. Case studies suggest that such factors may affect service delivery performance (Ferrinho and Van Lerberghe, 2000, 2002; Ferrinho et al., 1998; Tandler, 1997). However, case studies cannot support formal tests of hypotheses and are not replicable. Recently, efforts to develop more rigorous measures of intrinsic motivation in the health sector have been undertaken, but the difficulties in measuring health worker performance have limited the usefulness of these efforts (Franco et al., 2002).

So, the primary hindrances to agent-level, quantitative analysis of public servant performance relate to the measurement of performance and intrinsic motivations and the lack of institutional variation across otherwise directly comparable contexts. The second of these problems can be addressed through random policy interventions. However, we find only one example in the literature that also overcomes the problem of measuring performance: Nagin et al. (2002) observe the work effort of employees at a call center before and after a reduction in monitoring and find that it only leads to a reduction in effort among employees who perceive their employer as uncaring and unfair.

Another way to circumvent these problems is to run laboratory experiments. Bjorn and Schulze (2000) develop a principal-agent game in which the agent is offered a bribe by an exogenous party. Bribery yields a higher pay-off for the agent but causes damage to the principal. They find that certain subject types (male economists) are more corruptible than others, while in general agents are no less corruptible if they are paid more. Abbink et al. (2002) developed and Abbink (2004; 2005) applied a game in which one actor can choose to bribe another and thereby induce him or her to increase both their earnings at a cost to other subjects. The bribee can choose to accept or reject the bribe. There is a fixed

probability of detection, whereby both players are disqualified from the game and lose any prospect of further earnings. Abbink finds that reciprocity can support bribery and that rotation of staff reduces corruption. The threat of disqualification significantly reduces corruption, while higher wages do not reduce corruption, although corruption may, nevertheless, be motivated by fairness considerations.

Most recently, Azfar and Nelson (2007) developed a game designed to explore the effect of a division of power between an attorney-general (the monitor) and a president (the agent) on presidential resource expropriation. Using US undergraduates as the role-playing subjects, they found that making attorney-generals separately answerable to the electorate increased their vigilance, while increasing the earnings of the president and the ease with which expropriated resources could be detected reduced expropriation. Our experiment is similar to Azfar and Nelson's. The key differences are that: instead of an elected president, we have a randomly selected public service deliverer as the agent; instead of an elected or appointed attorney-general, we have an elected or randomly selected public servant as the monitor; we remove the political frame and variably replace it with neutral language and a frame that refers to a public health service delivery scenario; our subjects are not US undergraduates; and we take steps to increase the accuracy with which the monitors' intentions are revealed by their behaviour within the context of the game. The first four differences increase the extent to which the game reflects the context in which we are ultimately interested and allows us to test a series of hypotheses relating to professional norms through the interaction effects between framing and identity. The fifth change leads to the identification of a monitoring effect that was undetected in the original experiment.

3. The Experimental Design

3.1 The Public Servant's Game

The Public Servant's Game involves eight players, who in different rounds may play the role of 'community member', 'public service provider' and 'monitor'. On arrival, the players are assigned participant numbers from 1 to 8 and are then invited to sit around a table in full view of one another.

[Table 1 about here]

Table 1 contains a step-by-step summary of the game. The game begins when one player is selected to be the service provider according to the roll of an eight-sided die. Then, in the randomly selected monitor treatment, the eight-sided die is rolled again to select the monitor.⁴ In the elected monitor treatment, die rolls are used to select two candidates for the position of monitor and the five players whose numbers have not come up in the die rolling become voting community members. They each cast a secret ballot by writing the participant number of the candidate they want to be their monitor on a piece of paper. The ballots are collected by the experimenter, the votes are tallied, and the winner announced and appointed as the monitor. The unsuccessful candidate becomes the sixth ordinary community member. The service provider sits at the head of the table and is told that his or her wage for the round will be 20 Birr in the low wage treatment or 60 Birr in the high wage treatment.⁵ The monitor sits to the left of the service provider and is told that he or she will receive 60 Birr some of which may be used to engage in monitoring. The community members sit in the remaining chairs.

The role assignments complete, the playing of the round begins. The service provider rolls a six-sided die behind a screen to determine how many *valuable tiles* the experimenter is to place on a flat pad in front of the service provider but still behind the screen and hence out of sight of the other players. The valuable tiles represent the resources – drugs, dressings, syringes, equipment, etc. – the health worker has at her disposal to serve the community. The experimenter then adds *worthless tiles* to the pad making the

⁴ If the service provider's number comes up, the die is rolled again.

⁵ One US Dollar is worth 8 Birr. 60 Birr represents 10 to 15 % of basic monthly earnings for a beginning public sector nurse. In the private sector a nurse can earn around 60 Birr (extra) per night shift.

total number of tiles on the pad up to either 10 or 18 depending on the treatment. This variation affects the ease with which the monitor can expose resources kept by the service provider instead of being used to serve the community. Valuable and worthless tiles are identical on one side, while on the other the former are marked with a '**B**' (short for 'Birr', the currency in Ethiopia) and the latter are marked with a '**0**'. At this point in the game, the tiles are placed on the pad marked side up.

The service provider then has to decide how many valuable tiles to try and keep and how many to distribute to the community. She indicates her decision by picking six tiles (all valuable, all worthless, or a mixture) off the pad and placing them in a bag for distribution to the community members. While doing this she is careful not to reveal the tile mix either in the bag or remaining on the pad to the other players. The keeping of more valuable tiles by the service provider may be interpreted as making less effort or as expropriating more resources for private use or illegal sale. The number of valuable tiles that the service provider keeps on the pad is our incentive compatible (inverse) measure of service provider performance.

Before the tiles in the bag are distributed to the community members, the monitor can if he chooses attempt to expose any valuable tiles that the service provider has kept. The tiles on the pad are turned over so that valuable and worthless tiles look alike. Then the screen is removed and the pad placed in front of the monitor and in view of all the players. The monitor first states how many tiles he wishes to turn over. He can turn between zero and four tiles. Five Birr is deducted from the 60 Birr he received on assuming the role in the current round for each tile he says he is going to turn. Having stated how many tiles he wishes to turn, he then turns the tiles of his choosing. The number of tiles that the monitor states that he wishes to turn is our incentive compatible measure of monitoring effort.

The nature of the turned tiles has implications for the service provider's earnings for the round and the assignment of roles in the next round. These implications are described below. However, first, we describe the last remaining action to be undertaken in the current round. The bag containing the tiles that the service provider allocated to the community is passed around the table so that each community member can blindly draw (and not replace) one tile. They learn whether their own tile is valuable or

worthless, but not what other community members draw.⁶ Each valuable tile is worth 60 Birr to the community member that draws it. Community members who draw worthless tiles earn nothing for the round.

To the service provider, each valuable tile that she keeps on the pad is worth 40 Birr *as long as no valuable tiles are turned over by the monitor.*⁷ If one or more valuable tiles are turned over by the monitor, *all* the valuable tiles that the service provider has kept on the pad become worthless.

Once all the community members have drawn a tile, that round is finished, and the next round begins. If the incumbent service provider has been exposed as keeping resources, the eight-sided die is used to select the service provider for the next round.⁸ If the service provider has *not* been exposed as keeping resources, a 14-sided spinner is used to select the service provider for the next round. Eight sides of this spinner are numbered and six are marked ‘incumbent’. If the spinner settles on either the incumbent service provider’s number or one of the ‘incumbent’ sides, the incumbent service provider remains in post. So, the probability that the service provider remains in post for another round is 0.5. If the spinner settles on another number then the corresponding player becomes the service provider for the next round.

The same procedure for appointing the service provider is followed in the random and elected monitor treatments. However, while in the random treatment the incumbent monitor can be selected to become the service provider, in the elected monitor treatment, she cannot. This is because, in the elected monitor treatment, the incumbent monitor always stands for re-election by the community. So, if the spinner settles on the incumbent monitor’s number, it is spun again.

⁶ Some information may be communicated through reactions to draws, but this does not affect the nature of the game. Even if the identity of *all* the drawn tiles becomes public, the communities’ knowledge of the service provider’s behavior is still limited as they do not know how many valuable tiles the service provider received at the outset.

⁷ Accordingly, underperformance by the service provider causes a dead weight loss to society of 33.3 percent if not exposed and 100 percent if exposed.

⁸ If the number of the incumbent service provider comes up, the die is rolled again.

Once the service provider for the next round has been appointed, attention turns to the appointment of the monitor. In the random monitor treatment the eight-sided die is rolled as before. In the elected monitor treatment a new candidate for the monitor role is selected using the eight-sided die. The incumbent monitor and this new candidate then run for election. The successful candidate becomes the monitor. Then, play progresses as before.

Each session involves one game comprising of twelve rounds. The participants are paid in accordance with the outcome of *one* of these twelve rounds. This pay round is randomly and openly determined in front of the participants at the end of the experiment, using a 12-sided die. This is explained to the participants before play begins. We based pay on a randomly selected round rather than cumulative earnings to minimize income effects. Payments are calculated with reference to the data collected during play and placed in envelopes for distribution. The participants sign to verify payment and fill out a questionnaire before being dismissed.

When the monitor is randomly selected the equilibrium for this game involves no monitoring by monitors and the keeping of all valuable tiles by the service providers. However, when monitors are elected, multiple equilibria involving positive levels of monitoring and the passing of some valuable tiles to the community members are possible.⁹

3.2 Treatments

In the experiment, we vary four factors by design in order to see how they affect the service providers' and monitors' behaviour and how service recipients use their votes when given the opportunity to elect the monitor. The first of these factors is whether the monitor is randomly selected or elected by the community. This allows us to test three hypotheses: that service receiving communities are more likely to re-elect monitors who work hard; that monitors who are held accountable by recipient

⁹ Roth (1995) argues that economic experiments typically serve one of three purposes: (i) to test hypotheses from theory; (ii) to search for empirical regularities; or (iii) "to whisper in the ears of Princes", where the game is used as a metaphor for a more complicated situation in real life, and where the analysis provides tools that facilitate and structure policy analysis and advice. The experiment described here aims to serve the third purpose and for that reason we are not disturbed by the absence of a single equilibrium in the elected monitor version of the game.

communities through elections put greater effort into monitoring; and that, when facing community-elected monitors, service providers perform better.

The second factor that we vary is the total number of tiles placed on the pad in front of the service provider. This in turn affects the ease with which the monitor can expose resources kept by the service provider. This allows us to test the hypothesis that when the activities of a service provider are more readily observed, they perform better.

The third factor that we vary is the wage of the service provider. When set at 60 Birr this wage is equal to the upper limit of what a recipient community member can receive and the maximum amount that a monitor can receive. When set at 20 Birr this wage is half the minimum amount that a monitor can receive and below the expected amount to be received by community members who are served by a service provider who keeps nothing. This allows us to test whether service providers receiving a higher wage perform better. In addition, it allows us to test whether, due to concerns about fairness, monitors put more or less effort into exposing underperformance by service providers who are receiving a higher wage.

Finally, in order to investigate the role of professional norms in determining behavior, we vary the way in which the game instructions are presented to the participants. Here, we build on the premise that, if professional norms affect individual behavior, they do so only when individuals find themselves in a context in which their professional identity appears to them to be salient.¹⁰ Consider, for example, a stockbroker who is also a father facing two decisions that are mathematically equivalent but different with respect whether he perceives each as a decision for himself as a stockbroker or as a father. If there are professional norms for stock brokers that dictate a response different from the response he would make as a father, he will make different choices depending on whether he sees his identity as a stockbroker or a father as most salient.

Our subjects are members of the nursing profession who will soon be working in the Ethiopian public service. So, if they have internalized any norms relating to their profession, by varying the extent

¹⁰ The concept of identity takes a central role in sociology and psychology and argues that people's notions on how to behave depend on the situation. For an overview of how this affects – and may enrich – standard economic analysis, see Akerlof and Kranton (2000).

to which they perceive their identities as ‘nurse’ or ‘public servant’ as salient within the context of the experiment, while holding everything else constant, we may be able to identify the effect of professional norms on behaviour. We do this by presenting the Public Servant Game framed as a public health service delivery situation to half of the sample and in abstract form, i.e., making no reference to the government or health service delivery, to the other half. The description of the game above closely resembles the framed version of the game. In the unframed version: the service provider is referred to as ‘player P’, and the community members are referred to as ‘player Cs’ and valuable tiles are never likened to resources.

We have two hypotheses relating to the effects of framing on service provider and monitoring behavior. As nurses and public servants, we might expect our subjects refer and/or adhere to professional norms relating to the use of public resources to serve the community. Thus, when the experiment is framed as a public health service delivery situation, they would expropriate less and monitor more. However, having already gained work experience in a public sector in which underperformance is commonplace, our subjects may associate positions in public service with discretion over the use of resources for either public or personal ends. Thus, when the game is framed, the service provider and monitor may expropriate more and monitor less.

Further, it may be that the framing will cause different types of subject to reference different professional norms. We are particularly interested in finding out whether increased exposure to the Ethiopian public health sector is associated with increased observance of a norm of good public service or whether trainee nurses are effectively being socialized into a culture of corruption. If those who have more experience of working in the public health sector monitor more (less) and expropriate less (more) in the framed version of the game, we shall take it as evidence that the former (latter) is taking place.

3.3 Sessions

We ran sixteen experimental sessions. In eight of these the game was presented in abstract form and in the remaining eight it was framed as above. In eight sessions the monitor was randomly selected

and in the other eight the monitor was elected. And in eight sessions the service provider was paid the lower wage and in the other eight he or she was paid the higher wage. Finally, observability, i.e., whether there were 10 or 18 tiles in total on the pad when it was placed in front of the health worker, was varied within each session. Either rounds one to six were played with high observability (10 tiles on the pad) and rounds seven to twelve were played with low observability (18 tiles on the pad) or vice-versa. The design ensured that the resulting sample of 192 game rounds was balanced with respect to all four treatment variables and that the treatment variations were orthogonal.

In every session, before the twelve rounds of the game, four practice rounds were played. During four pilot sessions these practice rounds were found to be the best way to ensure that subjects fully understood the game before engaging in the incentive compatible rounds from which we draw our data for analysis.¹¹

4. The Experimental Context

Before turning to the results it is useful to describe the context in which the experiment was conducted and how that related to our experimental design and choice of subjects.

The scenario that the experiment aims to mimic is commonplace throughout the developing world: a large proportion of health care professionals are working alone or in teams of two or three in clinics often located in extremely remote areas; they are underpaid and poorly resourced; if they are monitored at all, it is by local government officials who generally lack the motivation, resources and skills needed to monitor effectively. We know that in contexts such as these health workers often

¹¹ We carried out four pilot sessions with four different groups of subjects. During these, a number of other simplifications were made to the game design, including switching from an increasing to a constant marginal price for monitoring, and the protocol was adjusted to maximize subject engagement with the physical elements of the game, i.e., the dice and the tiles. While, due to time constraints, we did not survey the subjects about their comprehension, throughout the pilot sessions we invited the subjects to ask questions and make comments and during the sessions proper the subjects were invited to raise their hand so that members of the research team could address any questions one to one. The questions asked were sensible and suggested a thorough understanding of what was going on. In the absence of an exit survey, uncertainty about subject understanding may remain. However, it is difficult to see how potential misunderstanding would bias the experimental results towards hypothesis acceptance.

expropriate drugs, dressings, syringes, and other consumable resources for use in private practices often run by themselves at times when they ought to be staffing public clinics or simply semi-privatize the public clinics by charging so called ‘side payments’ for services rendered (Jafré and Olivier de Sardan (2003), World Bank (2004), Transparency International (2006)).

Evidence suggests that this type of scenario is particularly commonplace in Ethiopia, where over 80 percent of the population dwells in variably remote rural areas. Ethiopia’s health outcome statistics and health service indicators are amongst the worst in the world (World Bank 2004). Comparable surveys of outpatients in Ethiopia, Bangladesh, and Trinidad and Tobago returned rates of patient satisfaction of 54 percent, 68 percent, and 74 percent respectively and have led to the conclusion that health worker performance should be a concern in all three countries, but especially in Ethiopia (Cueto 2005). Interviews with Ethiopian health workers confirm that they frequently expropriate drugs and other consumables, charge for health care provision when no charge should be made, and indulge in widespread absenteeism (Lindelov and Serneels 2006; Serneels et al. 2008). And surveys of student nurses indicate that many enter the sector already anticipating opportunities for misconduct (Serneels et al. 2005).

There is also evidence indicating that many of these problems are linked to flaws in institutional design, dysfunctional workplace norms and low levels of intrinsic motivation (Serneels and Lievens 2008). Earnings in the public sector, for example, remain very uniform and are not linked to performance. Further, even serious misconduct appears not to be punished, rendering the expected cost of detection low if not zero (Serneels et al. (2008)). And even health workers themselves acknowledge that misconduct is becoming more widespread (Lindelov and Serneels 2006) possibly in step with the apparent general increase in corruption in Ethiopia reported by Transparency International.¹²

¹² Ethiopia’s Corruption Perception Index and ranking deteriorated between 2002 and 2003. Ethiopia was ranked 61 out of 102 countries in 2002, the year corruption was first measured in Ethiopia by Transparency International. In 2003, the year of the experiment, it was ranked 94 out of 133.

In response to this widespread institutional failure in the health sector, policy makers in Ethiopia are experimenting with a shift away from monitoring, in principle, by district governments towards monitoring by local peasant and urban dwellers' associations, who are appointed by the community.

There are three ways in which a natural context can be brought to bear on an experiment. First, the game can be designed so that the incentive structures reflect key elements in the natural context. So, for example, in our design the elected monitor represents an officer of the local peasants association charged with the duty of monitoring on behalf of the community, while the randomly selected monitor represents a relatively disinterested district government officer charged with the same duty. Second, the experiment can be framed in a way that invokes the context in the minds of the subjects – hence, our references to the health worker, community, clinic, and drugs and dressing in the framed version of the game. And third, the experimental subjects can be selected specifically because they are familiar with the context and may, thus, reveal behavioural tendencies acquired within that context by the way they play the game.¹³ Arguably, the ideal subjects for our experiment would have been in-service health workers in rural postings. However, due to practical constraints and a concern that a subject's ability to understand the decisions they were being asked to make might decline over time away from the classroom, we chose to engage nursing students instead, safe in the knowledge that most Ethiopian nursing students have direct experience of the health sector either at the time of entering their program of study or shortly after through their training.

[Table 2 about here]

The subjects were sampled by, first, randomly selecting one government, one NGO, and one private nursing school and then randomly selecting students from within each school.

¹³ Insights relating to such behavioral tendencies may then be of value when seeking to understand outcomes in the natural world (see Levitt and List 2006 and Cardenas and Carpenter 2008).

Students could refuse to participate, but refusals were rare and there was no apparent pattern in the reasons given.¹⁴ Table 2 provides descriptive statistics relating to the subjects.

Out of the 144 subjects, 56% were *Female*. Their average age was 21 years, although ages ranged from 18 to 38. They had been in *Training* as nurses for between one and four years, the average being one year and ten months. Prior to entering nursing college, they had completed between 12 and 19 years of *Education*, the average being less than 14 years. Almost all of them (94 percent) were *Full time* students and 7 percent had paid *Work experience* in the health sector over and above experience gained during their training.

5. Results

In this section we discuss the results relating to each of the decision variables in the experiment. First, we look at the performance of service providers. Then, we analyze the behaviour of monitors. Finally, we analyse the behaviour of service recipients.

5.1 Service Provider Performance

A participant assuming the role of service provider in a round of the game can earn between 20 and 260 Birr in the low wage treatment and 60 and 300 Birr in the high wage treatment. So, being a service provider is potentially lucrative. Although, in the low wage treatment, it is only more lucrative than being a monitor if one keeps tiles and is undetected. On average, service providers earned 67.90 and 84.20 in the low and high wage treatments respectively.

In the randomly selected monitor treatment we observed behavior consistent with the standard assumptions of selfish money maximization, i.e., the retaining of all the valuable tiles, in one third of the

¹⁴ By paying particular attention to the representativeness of our sample, we increase the external validity of the insights we generate relating to behavioural tendencies in the natural world (see Harrison and List 2004),

rounds played. In the remainder, either expected monitoring or other regarding preferences caused the service providers to pass at least some tile on to the community.

[Table 3 about here]

[Figure 1 about here]

Table 3 indicates that across all treatments the mean number of tiles retained was 1.61 and that this number varies markedly across treatments. Figure 1 displays the cumulative distributions of the numbers of tiles retained by service providers under the different treatments. From Table 3 and Figure 1a we see that when the monitors were randomly selected the service providers retained more tiles than when the monitors were elected.¹⁵ The means are 1.96 and 1.27 tiles respectively and, according to Somers' D test in which we account for the non-independence of observations within each session by clustering, the difference in the distributions is significant (1 percent level). Table 3 and Figure 1b indicates that observability also had an impact. When observability was high service providers retained an average of 1.36 tiles compared to an average of 1.86 tiles when observability was low and, according to a Somers' D test in which we account for within-session non-independence, the difference in the distributions is significant (0.5 percent level). Table 3 and Figure 1c also indicate that when service providers were paid more they retained fewer tiles. The means for lower and higher paid service providers are 1.90 and 1.33 tiles respectively and, once again, according to a Somers' D test, the difference in the distributions is significant (3 percent level). Finally, Table 3 and Figure 1d indicate that mean numbers of tiles retained in the abstract and framed treatments did not appear to differ.

[Table 4 about here]

To check the robustness of these treatment effects, we run a series of regressions that take the number of tiles retained by the public servant as the dependent variable, the four treatment indicators as the independent variables of interest, and as control variables (i) the number of valuable tiles received by

¹⁵ Azfar and Nelson did not observe such an effect, possibly because the election of the presidents themselves had a strong direct effect.

the service provider, (ii) and four subject characteristics, namely sex, age, years of education, and years of training as a nurse to date;¹⁶ and (iii) the round (to control for the possibility that participants become more or less inclined to expropriate as play progresses). Regression analysis also facilitates an investigation into whether the framing of the game affected different types of subjects in different ways. The regressions, all of which are linear with errors adjusted to take account of the non-independence within sessions by clustering, are presented in Table 4.

The regression in column 1 contains only the treatment dummy variables, the resources available to the service provider (number of valuable tiles determined by the die roll), and the round identifier. We report no regressions containing interactions between treatment dummies because none were found to be significant. In column 2 the four subject characteristics are included and in column 3 an interaction term between the subjects' age and *Framed* is also included. Interactions between other subject characteristics and *Framed* were found not to be significant.

The results confirm that service providers facing elected monitors retained significantly fewer tiles (between 0.67 and 0.71 fewer tiles) and that when observability was high the service providers also retained significantly fewer tiles (between 0.42 and 0.51 fewer tiles). The results also indicate that service providers retain more tiles when more were made available to them. The regressions in the first and second columns indicate that service providers who received a wage of 60 Birr rather than 20 Birr retained significantly fewer tiles. However, in column three, where the subjects' characteristics are more thoroughly controlled for, the effect of the change in wage is insignificant. The regression in column 2 suggests that resource expropriating behavior does not vary with subject characteristics. However, the results in column 3 indicate that the age of a subject affected their behavior in different ways depending on whether the session was framed or not.

In section 2 we hypothesised that subjects who have experienced greater exposure to the reality of the Ethiopian public health service are more likely to perceive the use of public resources to serve

¹⁶ Too few were part time and too few had previous work experience in the health sector to expect the inclusion of these variables to yield reliable coefficient estimates.

selfish ends as normal and acceptable and, so, expropriate more when the game is framed. To test this hypothesis we tried including interaction terms between the subjects' training to date and the framed treatment dummy and the subjects' age and the framed treatment dummy variable. The former was insignificant, while, as column 3 in Table 4 indicates, the latter was significant and rendered the coefficient on the framed treatment dummy significant as well. Based on the estimated coefficients in this model, an eighteen year old subject would keep just under half a tile less when the game is framed, whereas a 38 year old subject would keep almost two tiles more.¹⁷ That age and not years training affects the subjects' response to framing suggests that it may not be exposure while in or being prepared for the role of health worker but exposure to corruption in health and more generally that matters. We appear to be observing idealism giving way to cynicism.¹⁸

5.2 Monitoring Performance

A participant assuming the role of monitor can earn between 40 and 60 Birr in that round. On average, monitor's earnings were 53.60 Birr per round, significantly greater than the corresponding figure of 17.30 Birr for community members. Being a monitor is lucrative, although not as lucrative as being an undetected, expropriating service provider.

The behavior of randomly selected monitors was consistent with the standard assumption of selfish money maximization, i.e., no monitoring, in just under half (46 percent) of the rounds played. In the remainder, either some form of other regarding preference or simple curiosity caused monitors to turn one or more tiles.

[Figure 2 about here]

¹⁷ Both of these predictions are significantly different from zero at the 10 percent level.

¹⁸ To further test the robustness of the results reported above, we repeated the analysis using a Tobit model that allows for left and right censoring to allow for the upper and lower limits on expropriation imposed by the design. The magnitude of the treatment effects was slightly larger, while the levels of significance were similar. The treatment effects reported in Table 4 can therefore be viewed as conservative. We also estimated the models using random effects by subject and these also yielded very similar results.

Table 5 indicates that the mean number of tiles turned was 1.28; and that this number varied markedly across treatments. Figure 2 displays the cumulative distributions for the numbers of tiles turned under the different treatments. From Table 5 and Figure 2a, we see that randomly selected monitors turned fewer tiles than elected monitors. The mean numbers of tiles turned were 0.72 and 1.84 respectively and, according to Somers' D test in which we account for the non-independence of observations within each session by clustering, the difference in the distributions is highly significant (0.05 percent level). In contrast, both Table 5 and Figure 2b suggest that there was no change in monitoring behavior associated with the change in observability. Both Table 5 and Figure 2c suggest that the monitors turned more tiles when the service providers were better paid. The mean numbers of tiles turned in the low and high wage treatments were 1.04 and 1.52 respectively. However, according to Somers' D test in which we account for within-session non-independence the difference in the distributions is not significant ($p=0.18$). To investigate whether this insignificance was due to the much larger and, in the bivariate analysis, uncontrolled for effect of the election treatment, we also conduct Somers' D tests on the sub-samples of monitoring decisions made by randomly selected and elected monitors separately. Thus, we find that the higher service provider wage is associated with significantly more monitoring when monitors are randomly selected. Finally, once again, framing appears to have no clear effect and the mean numbers of tiles turned in the abstract and framed treatments are very similar at 1.30 and 1.26 respectively.

[Table 5 about here]

Table 6 contains a series of regressions each taking the number of tiles turned as the dependent variable. All of the regressions are linear with standard errors adjusted to take account of the non-independence of errors within sessions. The regression in column 1 contains only the treatment dummy variables, the resources available to the service provider, and the round indicator. The regression in the column 2 also contains the one interaction between treatment dummies that we found to be significant. In column 3, the four subject characteristics are introduced and in column 4 interaction terms between

framed (F), the dummy capturing whether the game was framed as a health service delivery situation, and the subjects' age and years of training are also included. Interaction terms between framed (F) and the subjects' sex and education were found not to be significant.

The positive effect on monitors' behavior of being elected remains significant. However, note the significant interaction term between Monitor elected (E) and Observability high (O) in columns 2 to 4. *Ceteris paribus*, when observability was low, an elected monitor turned between 0.57 and 0.84 tiles more than a randomly selected one and when observability was high, they turned between 1.22 and 1.41 tiles more. Note also that in column 3, the observability high treatment dummy bears a weakly significant negative coefficient, suggesting that increased observability caused randomly selected monitors to turn fewer tiles. This difference in the effect of observability on monitoring by elected and randomly selected monitors is consistent with the former believing that exposing expropriation might improve the chances of re-election. In contrast to the Somers' D test result reported above, the regressions indicate that monitors turned more tiles when the service provider earned a higher wage. *Ceteris paribus*, if the service provider received 60 Birr instead of 20 Birr, the monitor turned between 0.48 and 0.96 tiles more.

[Table 6 about here]

All of the regressions indicate that monitoring behavior was insensitive to both the resources available to the service provider (which of course the monitor did not know) and the number of rounds that had already been played. However, the regressions in columns 3 and 4 indicate that more educated subjects turned more tiles: an additional year of education was associated with the turning of between 0.32 and 0.35 tiles more.

Finally, in column 4, the interactions between *Framed* and *Age* and *Framed* and *Training* as well as the uninteracted variables, *Framed*, *Age*, and *Training* are all significant. The resulting coefficients indicate that younger subjects and subjects who have received less training to date monitor more when the game is framed as a public health service provision situation, whereas older subjects and subjects who have received more training to date monitor less when the game is framed. Based on the estimated

coefficients and standard errors, an eighteen year old subject who has received one year of training to date would turn 1.26 more tiles when the game is framed and this predicted difference in behaviour between the framed and unframed games is significantly different from zero at the 0.1 percent level. In contrast, the predicted monitoring behaviour of 26 to 34 year old subjects with one year of training is statistically indistinguishable between framed and unframed games. By the time they have completed two years of training, the predicted monitoring behaviour of subjects aged 19 to 23 is statistically indistinguishable between framed and unframed games, while 24 year olds turn half a tile less and 34 year olds turn 1.6 tiles less in the framed game as compared to the unframed game. And after four years of training, even a 21 year old subject would turn over 2 tiles less in the framed as compared to the unframed game.¹⁹

5.3 Community Voting Behavior

A participant assuming the role of community member in a round of the game earns either 0 or 60 Birr in that round. The more valuable tiles there are in the bag the greater each community member's chance of earning 60 Birr. The only way in which a community member can affect his earnings is by electing a monitor that will look after his interests. During the experiment a total of 96 elections were held (12 in each of eight sessions). In 88 of these, one of the candidates was an incumbent, i.e., they had been monitor during the preceding round.

[Figure 3 about here]

Figure 3 indicates the average number of votes received by incumbent monitors depending on how many tiles they turned in the preceding round of the game and Table 7 reports the results of a regression of the number of votes the incumbent monitor received on the number of tiles they turned. Both suggest that monitors who made more effort, i.e., turned more tiles, received more votes, and the

¹⁹ To further test the robustness of these results, we also repeated the analysis using a Tobit model that allows for left and right censoring and obtained slightly larger treatment effects and similar levels of significance, indicating that the reported treatment effects can be viewed as conservative. Estimating the models using random effects by subject also returns very similar results.

result is highly significant throughout. The second and third columns of this table indicate that neither the success of the monitor at exposing service providers who were keeping resources nor any of the other treatment variables, like the wage level or the level of observability, affected voting behavior.

[Table 7 about here]

6. Conclusions

From the literature we identified four possible determinants of service provider performance: whether service recipients have the power to hold those charged with the duty of monitoring service providers accountable, the ease with which the performance of service providers can be monitored, service provider pay, and adherence to professional norms. We performed an experiment designed to test a series of hypotheses relating to these determinants. Now we review our results relating to each of the hypotheses.

In accordance with economic theories, we find that service providers perform better when their monitors have been elected by the service recipients, and that elected monitors put greater effort into monitoring. We also find that service providers perform better when observability is higher; and that community members re-elect monitors who put more effort into exposing expropriation.

Our results provide only weak evidence that public servants who receive a higher wage expropriate less. And, even when an effect is identified, in accordance with the findings of Van Rijckeghem and Weder (2001), it is small. Increasing the wage from 20 Birr to 60 Birr, i.e., a 200% increase in the wage, leads to less than a 30 percent reduction in resource expropriation. Unlike studies using cross-country data, we were able to distinguish between the effect of the wage increase on the service provider and the effect of the same increase on the effort of monitors. We found that monitors put more effort into exposing underperformance by public servants who receive a higher wage. This corroborates Abbink's (2004) observation that notions of distributive fairness may affect behavioral responses toward changes in relative pay.

Our experimental design also allowed us to tentatively explore the effects of professional norms and workplace culture on behaviour. By running both framed and unframed versions of the game we endeavoured to vary the apparent salience of professional norms associated with nursing and public service in Ethiopia and thereby identify their effects on behavior. Framing had no observable effect on either mean monitoring or mean expropriating behavior. However, further investigation revealed that the behaviour of young subjects and subjects early in their training was positively affected by the framing, i.e., they expropriated less and monitored more, while that of older subjects and subjects nearing the completion of their training was negatively affected by the framing, i.e., they expropriated more and monitored less. These results suggest that subjects who have experienced less exposure to the reality of the Ethiopian public health service may be adhering to positive professional norms of public service within a caring profession, but that exposure may lead to an erosion of these norms as the use of public resources to serve selfish ends is increasingly perceived as normal and acceptable. This finding suggests that new recruits into the Ethiopian public health service are being corrupted through either some sort of social learning or secondary socialization process. However, it also suggests that a positive norm of public service in the caring professions still exists.

References

- Abbink, K. 2004. Staff Rotation as an Anti-Corruption Policy: An Experimental Study. *European Journal of Political Economy*, 20, 887-906.
- Abbink, K. 2005. Fair Salaries and the Moral Costs of Corruption. In Kokinov, Boicho (Eds), *Advances in Cognitive Economics*. NBU Press Sofia.
- Abbink, K., Irlenbusch, B., Renner, E. 2002. An experimental bribery game. *Journal of Law, Economics and Organization* 18, 428-454.
- Ablo, E.Y., Reinikka, R. 1998. Do budgets really matter? Evidence from public spending on education and health in Uganda. Policy Research Working Paper 1926. Washington, DC. World Bank.

- Andvig, J.C., Fjeldstad, O.H. 2001. Corruption: A review of contemporary research. CMI report R 2001:7: Chr. Michelsen Institute.
- Akerlof G., Kranton, R. 2000. Economics and identity. *Quarterly Journal of Economics* 115, 715-53.
- Arrow, K.J. 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* 53, 941-73.
- Azfar, O., Nelson, W. R. J. 2007. Transparency, wages, and the separation of powers: An experimental analysis of the causes of corruption. *Public Choice* 130, 471-493.
- Banfield, E.C. 1979. Corruption as a feature of government organization. In Ekpo M. (Ed.), *Corruption In Sub-Saharan Africa: Towards A Search For Causes And Consequences*. Washington DC: University Press of America.
- Bardhan, P. 1997. Corruption and development: A review of issues. *Journal of Economic Literature* 25 , 1320-1346.
- Becker, G.S., Stigler, G.J. 1974. Law enforcement, malfeasance, and the compensation of enforcers. *Journal of Legal Studies* 3, 1-18.
- Besley, T., McLaren, J. 1993. Taxes and bribery: the role of wage incentives. *The Economic Journal* 103 , 119-141.
- Björn F., Schulze G.G. 2000. Does economics make citizens corrupt? *Journal of Economic Behavior and Organization* 43, 101-113.
- Cardenas, J-C, Carpenter, J. 2008. Behavioral development economics: Lessons from field labs in the developing world. *Journal of Development Studies* 44, 337-364.
- Chand, S.K., Moene, K.O. 1999. Controlling fiscal corruption. *World Development* 27, 1129-1140.
- Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge, MA. Harvard University Press.
- Cueto M, 2005, The promise of primary health care. *Bulletin of the World Health Organization* 83, 322.
- Deci, E.L. 1975. *Intrinsic Motivation*. New York. Plenum Press
- Dixit, A. 1997. Power of incentives in private versus public organisations. *American Economic Review Papers and Proceedings* 87, 378-382.

- Dixit, A. 2001. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources* 37, 696-727.
- Ferrinho, P., Van Lerberghe, W., Julien, M.R., Fresta, E., Gomes, A., Dias, F., Goncalves, A., Backstrom, B. 1998. How and why public sector doctors engage in private practice in Portuguese-speaking African countries. *Health Policy Plan* 13, 332-338.
- Ferrinho, P., Van Lerberghe, W. 2000. Providing health care under adverse conditions: Health personnel performance and individual coping strategies. In: Lerberghe, W.V., Kegels, G. Vand den.Brouwere (Eds.), *Studies in Health Services Organisation and Policy*, Vol. 16. Antwerp: ITGPress.
- Ferrinho, P., Van Lerberghe, W. 2002. Managing health professionals in the context of limited resource: A fine line between corruption and the need for moonlighting..see www-wds.worldbank.org
- Fisman, R., Gatti, R. 2000. Decentralization and corruption. Policy Research Working Paper No. 2290: World Bank.
- Franco, L.M., Bennett, S., Kanfer, R. 2002. Health sector reform and public sector health worker motivation: a conceptual framework. *Social Science and Medicine* 54, 1255-1266.
- Frank, R.G. 2004. Behavioral economics and health economics. NBER Working Paper 10881.
- Frey, B.S. 1997. On the relationship between intrinsic and extrinsic work motivation. *International Journal of Industrial Organization* 15, 427-439.
- Glaeser, E.L., Shleifer, A. 2001. Not-for-profit entrepreneurs. *Journal of Public Economics* 81, 99-115.
- Harrison G.W., List J.A. 2004. Field experiments. *Journal of Economic Literature* 42, 1009-1055.
- Khalil, F., Lawarree, J. 1995. Collusive auditors. *American Economic Review* 85, 442-446.
- Klitgaard, R. 1988. Controlling corruption. Berkeley: University of California Press
- Jaffre Y., Olivier de Sardan J.-P. 2003. Une Médecine Inhospitière. Les Difficiles Relations entre Soignants et Soignes Dans Cinq Capitales d'Afrique de l'Ouest. Paris - Marseille, APAD-Karthale.
- Le Grand, J. 2003. Motivation, Agency, and Public Policy: Of Knights and Knaves, Pawns and Queens. Oxford: Oxford University Press

- Levitt, S. D., List, J. A. 2006. What do laboratory experiments measuring social preferences tell us about the real world? *Journal of Economic Perspectives* 21, 153-174.
- Lindelow, M., Serneels, P., 2006. The performance of health workers in Ethiopia: results from qualitative research. *Social Science and Medicine* 62, 2225-2235.
- Lindenberg, S. 2001. Intrinsic Motivation in a New Light. *Kyklos* 54, 317-342.
- Manor, J. 1999. *The Political Economy of Democratic Decentralization*. Washington DC: The World Bank
- Mookherjee, D., Png, I.P.L. 1995. Corruptible law enforcers: how should they be compensated? *Economic Journal* 105, 149-159.
- Nagin, D.S., Rebitzer, J.B., Sanders, S., Taylor, L.J. 2002. Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review* 92, 850-873.
- Ostrom, E., Walker, J., Gardener, R., 1992. Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86, 404-417.
- Pauly, M.V. 1987. Nonprofit firms in medical markets. *American Economic Review* 77, 257-262.
- Prud'homme R., 1995. The dangers of decentralization. *World Bank Research Observer* 10, 201-220.
- Rauch, J.E., Evans, P.B. 2000. Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of Public Economics* 75, 49-71.
- Rose-Ackerman, S. 1978. *The Political Economy Of Corruption*. New York: Academic Press
- Roth, A.E., 1995, Introduction to Experimental Analysis. In: Kagel J.A. and Roth A.E. (Eds.). 1995 *Handbook for Experimental Economics*, Princeton NJ. Princeton University Press.
- Serneels P., Lindelow, M., Montalvo J.G., Barr A., 2005, An honourable calling? unpublished report forthcoming as book, The World Bank.

- Serneels, P., Lievens T., 2008, Institutions for health care delivery. A formal exploration of what matters to health workers. Evidence from Rwanda. Working paper 2008-29, Center for the Study of African Economics, Economics Department, University of Oxford.
- Serneels P., Serra D., Lindelow M., Montalvo J.G., and Barr A., 2008, Discovering the real world. Health workers' early work experience and career choice. Working Paper, Ethiopia, The World Bank.
- Serneels P., Lindelow M. and Lievens T., 2008, Qualitative research to prepare quantitative analysis. Absenteeism among health workers in two African countries. In Amin S., Das J. and Goldstein M., 2008, Are you being served. New tools for measuring service delivery. The World Bank.
- Svensson, J. 2003. Who must pay bribes and how much? Evidence from a cross-section of firms. Quarterly Journal of Economics 118, 207-230.
- Transparency International, 2006, Global Corruption Report, Corruption and Health
- Tendler, J. 1997. Good Government in the Tropics. Baltimore. London. Johns Hopkins University Press
- Treisman, D. 2000. The causes of corruption: a cross-national study. Journal of Public Economics 76, 399-457.
- Van Rijckeghem, C., Weder, B. 2001. Bureaucratic corruption and the rate of temptation: Do wages in the civil service affect corruption, and by how much? Journal of Development Economics 65, 307-331.
- Wilson, J.Q. 1989. Bureaucracy: What Government Agencies Do and Why They Do It. New York Basic Books.
- World Bank, 2004, World Development Report, Making Services Work for the Poor. The World Bank.

FIGURES AND TABLES

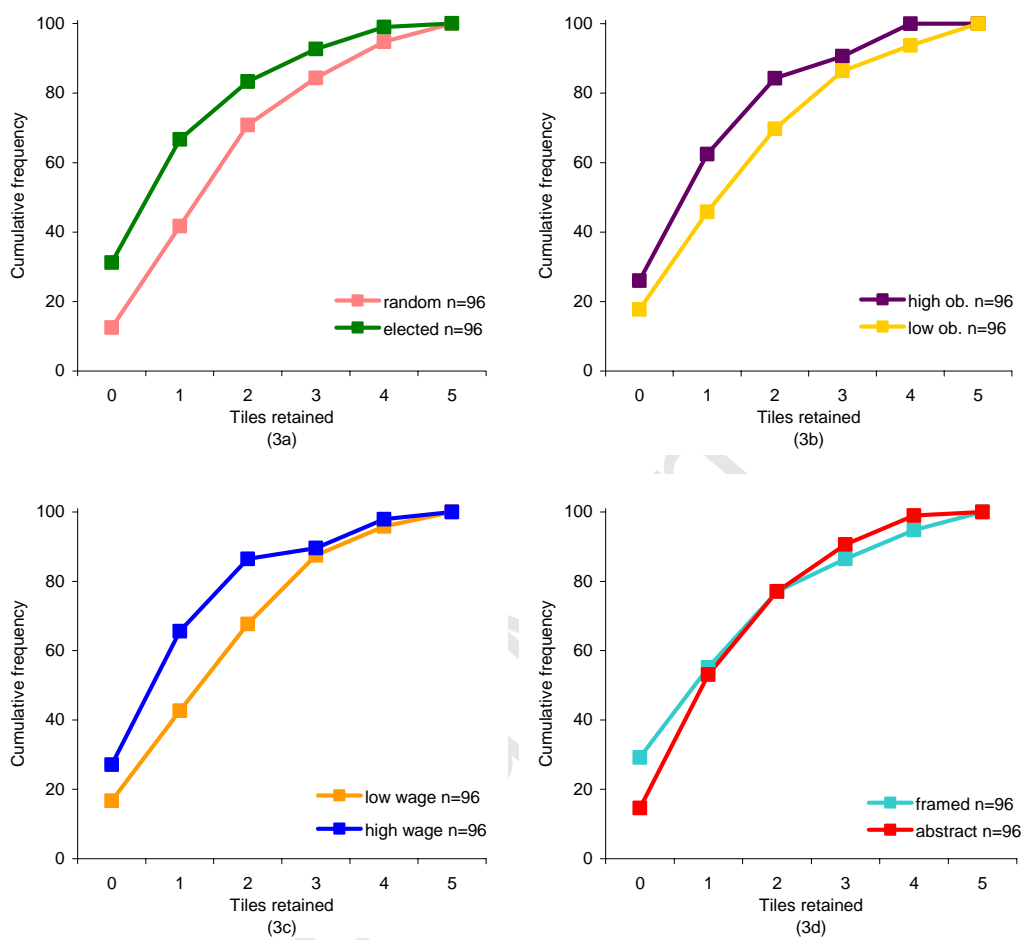
FIGURE 1: RESOURCES RETAINED BY SERVICE PROVIDER UNDER DIFFERENT TREATMENTS

FIGURE 2: MONITORING BEHAVIOR UNDER DIFFERENT TREATMENTS

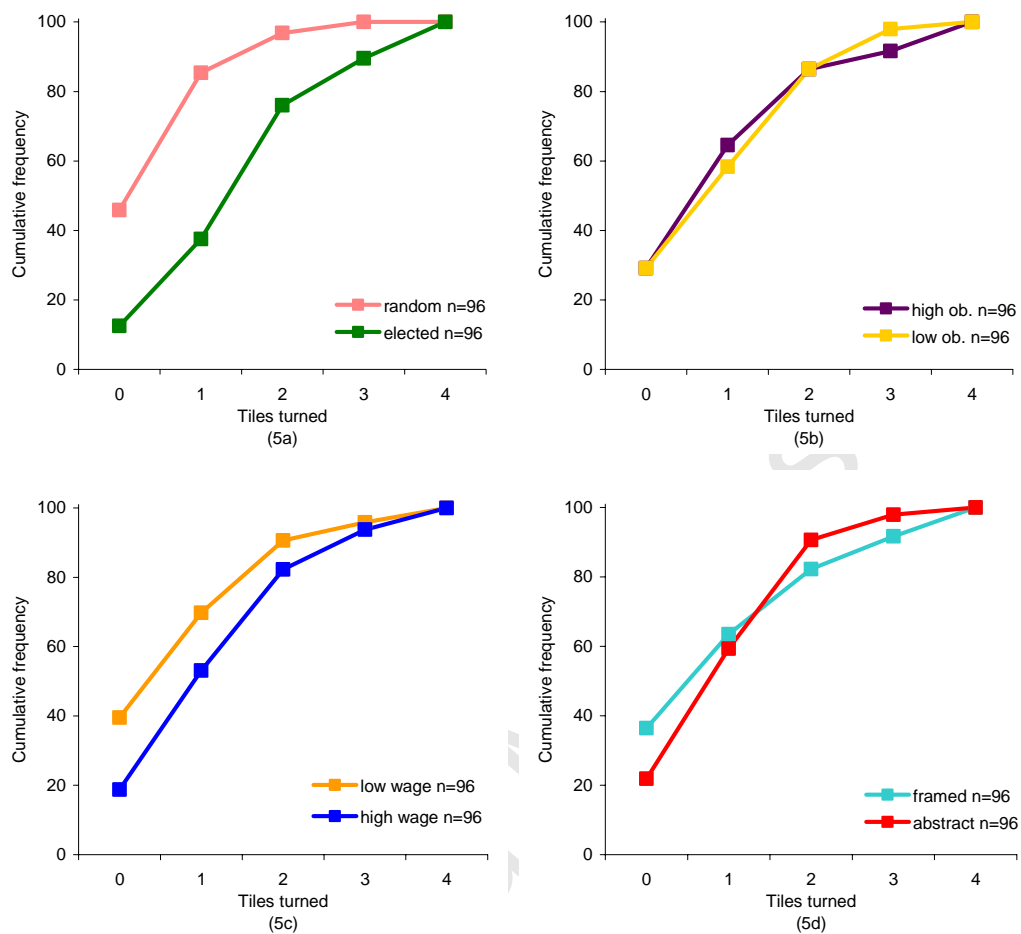


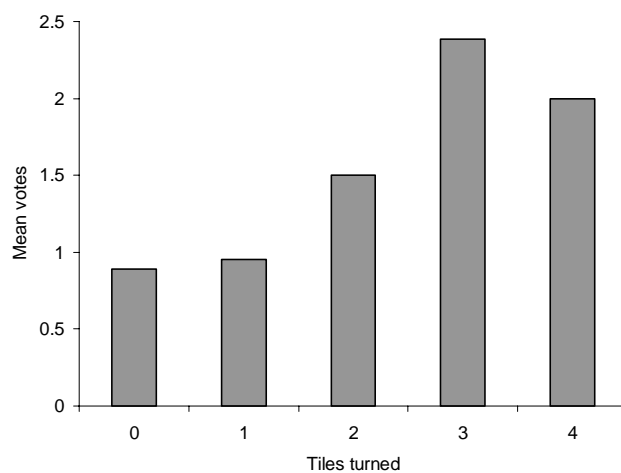
FIGURE 3: VOTES AND MONITORING EFFORT

TABLE 1: THE STRUCTURE OF THE GAME

Monitor is randomly selected	Monitor is a community representative
1. Service provider selected using 8-sided die	1. Service provider selected using 8-sided die
2. Monitor selected using 8-sided die	2. Two candidates for monitor selected using 8-sided die and community votes to elect one
3. Service provider rolls a 6-sided die to determine the number of valuable tiles (resources) she gets	3. Service provider rolls a 6-sided die to determine the number of valuable tiles (resources) she gets
4. Service provider decides how many valuable tiles to keep	4. Service provider decides how many valuable tiles to keep
5. Monitor states how many tiles (0 to 4) he is going to turn in order to try and expose a valuable tile that the service provider has kept. Then he turns	5. Monitor states how many tiles (0 to 4) he is going to turn in order to try and expose a valuable tile that the service provider has kept. Then he turns
6. If the monitor exposes a valuable tile, the health worker loses all the valuable tiles he has kept, and is excluded from being service provider in the next round	6. If the monitor exposes a valuable tile, the health worker loses all the valuable tiles he has kept, and is excluded from being service provider in the next round
7. Six tiles including all the valuable tiles that the health worker has not kept are randomly distributed to the six community members	7. Six tiles including all the valuable tiles that the health worker has not kept are randomly distributed to the six community members
8. If incumbent service provider is exposed, start again at 1. If incumbent service provider's numbers come up, roll again. If incumbent service provider not exposed, select service provider for next round using 14-sided spinner - incumbent remains if own number or 'incumbent' comes up. Continue at 3.	8. If incumbent service provider is exposed, roll 8-sided die to select new service provider. If incumbent service provider's or monitor's numbers come up, roll again. If incumbent service provider not exposed, select service provider for next round using 14-sided spinner - incumbent remains if own number or 'incumbent' comes up. If incumbent monitor's number comes up, spin again
	9. Incumbent monitor faces re-election. Roll 8-sided die to select the other candidate. Continue at 2.

TABLE 2: SOCIO-ECONOMIC CHARACTERISTICS OF SUBJECTS

	Means and proportions	Standard deviations	Range
All participants (Observations=participants=144)			
Female (%)	56.25		
Age (years)	21.34	3.01	18-38
Education (years in school)	13.66	0.99	12-19
Training (years on course)	1.85	0.72	1-4
Work experience (% with)	6.94		
Full time (%)	94.44		
Took healthworker role (%)	61.81		
Took monitor role (%)	68.75		
Characteristics of 'health workers' (Observations=rounds of play=192)			
Female (%)	53.65		
Age (years)	21.39	2.91	18-34
Education (years in school)	13.55	0.87	12-16
Training (years on course)	1.77	0.70	1-3
With work experience (%)	6.25		
Full time (%)	91.50		
Characteristics of 'monitors' (Observations=rounds of play=192)			
Female (%)	51.56		
Age (years)	21.37	2.77	18-34
Education (years in school)	13.54	0.82	12-16
Training (years on course)	1.78	0.71	1-4
With work experience (%)	5.21		
Full time (%)	94.79		

TABLE 3: RESOURCES RETAINED BY SERVICE PROVIDER

Tiles retained	n	Mode	Mean	Standard dev.	P-value for Somers' D test
Full sample	192	1	1.61	1.34	
Monitor randomly selected	96	1-2	1.96	1.35	0.60%
Monitor elected	96	1	1.27	1.24	
Observability high	96	1	1.36	1.21	0.20%
Observability low	96	1	1.86	1.42	
Wage low	96	1	1.90	1.36	2.90%
Wage high	96	1	1.33	1.26	
Abstract	96	1	1.66	1.19	58.50%
Framed	96	0	1.57	1.47	

TABLE 4: REGRESSION ANALYSIS OF TILES RETAINED BY SERVICE PROVIDERS

Dependent variable = tiles retained

	1	2	3
Monitor elected	-0.6832 *** [0.1767]	-0.6749 *** [0.2085]	-0.7088 *** [0.2082]
Observability high	-0.5055 *** [0.1154]	-0.4741 *** [0.1355]	-0.4242 ** [0.1456]
Wage high	-0.4893 ** [0.1746]	-0.4531 * [0.2297]	-0.4036 [0.2327]
Framed	-0.0274 [0.1772]	-0.0256 [0.1878]	-2.6466 * [1.2532]
Round (continuous)	0.0352 * [0.0199]	0.0319 [0.0215]	0.0353 [0.0239]
Resource	0.4132 *** [0.0372]	0.4139 *** [0.0365]	0.4098 *** [0.0376]
Female		-0.2292 [0.2010]	-0.2473 [0.2074]
Age		-0.0027 [0.0370]	-0.0855 * [0.0468]
Education		-0.0420 [0.0806]	-0.0470 [0.0879]
Training		0.0818 [0.1809]	0.0641 [0.1871]
Framed x Age			0.1217 * [0.0603]
Constant	0.8587 *** [0.1972]	1.4448 [1.4360]	3.3080 *** [0.9616]
Observations	192	192	192
R-squared	0.4279	0.4354	0.4492

Notes: Standard errors have been adjusted to take account of non-independence of errors within sessions. *** significant at 1% level; **significant at 5% level; * significant at 10% level.

TABLE 5: MONITORING BEHAVIOR

Tiles turned	n	Mode	Mean	Standard dev.	P-value for Somers' D test
Full sample	192	1	1.28	1.13	
Monitor randomly selected	96	0	0.72	0.79	0.05%
Monitor elected	96	2	1.84	1.14	
Observability high	96	1	1.28	1.19	67.70%
Observability low	96	0-1	1.28	1.07	
Wage low	96	1	1.04	1.09	17.90%
Wage high	96	2	1.52	1.11	
Abstract	96	1	1.30	0.96	71.10%
Framed	96	0	1.26	1.28	

TABLE 6: REGRESSION ANALYSIS OF MONITORING BEHAVIOR

Dependent variable = tiles turned

	1	2	3	4
Monitor elected (E)	1.1249 *** [0.2173]	0.8388 *** [0.2761]	0.5738 * [0.2914]	0.6903 *** [0.2270]
Observability high (O)	0.0002 [0.1367]	-0.2714 [0.1615]	-0.3197 * [0.1731]	-0.3074 [0.1961]
Wage high	0.4782 ** [0.2136]	0.4785 ** [0.2140]	0.5807 ** [0.2199]	0.9643 *** [0.1810]
Framed	-0.0424 [0.2185]	-0.0422 [0.2189]	0.0516 [0.2123]	4.3712 *** [0.6841]
Round (continuous)	-0.0006 [0.0174]	-0.0101 [0.0164]	-0.0207 [0.0163]	-0.0234 [0.0166]
Resource	-0.0054 [0.0353]	-0.0037 [0.0353]	-0.0170 [0.0396]	-0.0118 [0.0279]
E x O		0.5723 ** [0.2465]	0.6442 ** [0.2452]	0.6923 ** [0.2608]
Female			0.3555 [0.2530]	0.3267 [0.2101]
Age			0.0019 [0.0272]	0.0595 ** [0.0261]
Education			0.3244 ** [0.1346]	0.3522 *** [0.1181]
Training			-0.1012 [0.1493]	0.3489 ** [0.1468]
Framed x Age				-0.1123 *** [0.0294]
Framed x Training				-1.0862 *** [0.1872]
Constant	0.5226 * [0.2626]	0.7144 ** [0.2959]	-3.569 * [1.8254]	-6.2931 *** [1.5302]
Observations	192	192	192	192
R-squared	0.296	0.3114	0.3650	0.4446

Notes: Standard errors have been adjusted to take account of non-independence of errors within sessions. *** significant at 1% level; **significant at 5% level; * significant at 10% level.

TABLE 7: REGRESSION ANALYSIS OF VOTES FOR INCUMBENT MONITORS

	1	2	3
Tiles turned in previous round	0.4076 *** [0.0775]	0.4166 *** [0.0930]	0.4352 *** [0.0729]
Resources exposed in previous round		-0.0354 [0.1840]	
Wage high			-0.3965 [0.2816]
Observability high			0.3408 [0.2458]
Framed			-0.0858 [0.2664]
Constant	0.7105 ** [0.2216]	0.7102 ** [0.2232]	0.7207 ** [0.2386]
Observations	88	88	88
R-squared	0.1286	0.1290	0.1730

Notes: Standard errors have been adjusted to take account of non-independence of errors within sessions. *** significant at 1% level; **significant at 5% level; * significant at 10% level.